When gesture does and does not follow speech in describing motion

Şeyda Özçalışkan

1. Introduction

Learning language involves not only mastery of grammar and vocabulary, but also appropriate use of spontaneous gestures that accompany speech. As such, gestures constitute part of what learners acquire in a new language and show variability across different languages (Gullberg, 1998; Stam, 2006). But how early does gesture take on the patterns of the language one speaks? A good way to approach this question is by studying the speech and gestures produced by advanced bilingual speakers of structurally different languages, and examine whether their gestures resemble the patterns of their first (L1) or second language (L2) when speaking L2. If acquisition of native-like gesture patterns is a by-product of acquisition of native-like speech patterns, then we would predict that advanced L2 learners of English would produce gestures that resemble the gestures of native English speakers. If, on the other hand, acquisition of native-like gesture patterns takes longer than acquisition of native-like speech patterns, then we would predict that advanced L2 speakers of English would follow L1 gesture patterns when speaking L2.

Gesture and speech form an integrated system both semantically and temporally in the adult speakers; 90% of the gestures adults produce co-occur with speech and conveys distinct but related information (McNeill, 1992, 2005). Adult gestures also show variability across different languages, suggesting the existence of gestural repertoires unique to particular languages (Kendon, 1986, 2004). The existence of such gestural repertoires, in turn, suggests that gestures should be treated as part of what learners need to acquire when learning a new language. But how early does this gestural repertoire show the language-specific patterns found in speech? One way to approach this question is to study the gestures that L2 speakers produce when describing a domain that shows strong cross-linguistic variability in native speakers’ speech and gesture, and to explore whether L2 speakers’ gestures resemble the gestures of native speakers speaking the same language.

Spatial motion, which shows wide variation as well as patterned regularities across the world’s languages, offers a unique arena within which L1 vs. L2 effects on gesture use can be examined. Previous research (Talmy, 1985, 2000) identified the ‘motion event’ as a basic building block of language and cognition, and proposed a set of motion components that are universal. For example, in expressing a simple motion scene, such as a boy running into a house, many languages make it possible to refer to the figure (boy) separately from the ground (house), to trace its path (in, out), or to comment on the manner with which he moves (crawling, running). However, at the same time, languages show systematic variation in how
they express each of these motion components. As proposed by Talmy, the world’s languages can be categorized into two in terms of their expression of spatial motion: In *satellite-framed languages* (of which English is one), manner of motion is typically expressed in the verb, while path of motion appears in a particle outside the verb (e.g., ‘The boy ran into the room’); whereas in *verb-framed languages* (of which Turkish is one), the verb usually encodes the path of motion, while manner information is optionally expressed with gerunds or adverbs outside the verb (e.g., ‘The boy entered the room [running]’).

Differences in the expression of path of motion (inside vs. outside the verb) have consequences for whether path or manner of motion is frequently mentioned by speakers of each language. The preference of English speakers to express path by particles leaves the main verb available for manner information. Since the main verb is the obligatory core of a sentence, this gives English speakers an easily codable linguistic option to convey manner. As an outcome, English speakers develop a habitual pattern of expressing manner (Slobin, 2000). In contrast, because Turkish speakers typically express path in the main verb, they have to use subordinate grammatical constructions, such as gerunds (e.g., ‘exit running’) or adverbs (e.g., ‘exit hastily’), to convey manner. Since such constructions add an extra processing load, Turkish speakers are less likely to mention manner in their expression of motion than English speakers (Slobin, 2004).

Previous work has provided evidence for the cross-linguistic differences in the expression of motion between adult English and Turkish speakers (Özçalıskan & Slobin, 1999, 2003; see also Özçalıskan, 2004, 2007, 2009). English speakers use a significantly larger number and variety of verbs indicating manner (e.g., ‘run’, ‘plunge’) than Turkish speakers, whereas Turkish speakers rely on a limited set of path verbs (e.g., ‘ascend’, ‘exit’) and use them at significantly greater rates than English speakers. Furthermore, in English, the frequent expression of manner in the verb has a spillover effect on the choice of other lexical items. English speakers use not only a greater rate and variety of manner verbs but also a greater number of other lexical items (adverbs, verb complements) that indicate manner, pointing to their greater awareness of this dimension as a salient conceptual category (Özçalıskan, 2004, 2005). Thus, earlier work on speech about motion shows that native English and Turkish speakers differ in their expression of spatial motion, particularly with respect to manner and path components.

However, speakers use not only words but also gestures when describing spatial motion. Earlier work has in fact shown that adult speakers of different languages gesture a great deal when talking about motion (e.g., Kita et al., 2001; Nunez & Sweetser, 2006). More importantly, these gestures show cross-linguistic differences, often mirroring the patterns found in speech. An earlier study of gestures accompanying motion descriptions in English and Turkish showed that English speakers typically synthesize manner and path components into a single gesture (e.g., circling hands while moving them forward as if rolling forward), while Turkish speakers produce separate gestures for manner and path—replicating the patterns found in their speech (Kita & Özyürek, 2003). These results make it clear that native adult speakers of English and Turkish differ not only in their speech about motion, but also in the gestures they produce while describing motion. The question we ask here is whether L2 speakers of English display the same gesture patterns as their native counterparts when describing motion.
Although there are now several cross-linguistic descriptions of gesture and speech production in native speakers, relatively little is known about the gestures of L2 speakers. Of the existing studies, some focused on the amount of gesture production and asked whether bilinguals used more or fewer gestures when speaking their stronger language (L1) as opposed to their weaker language (L2). This earlier work focused on two predominant gesture types, namely deictics (gestures that indicate referents; e.g., pointing at a ball) and iconics (symbolic gestures that convey aspects of intended referent; e.g., holding cupped hands in air as if it is a ball, moving empty hand forward forcefully as if throwing), and provided evidence for both possibilities. Research on deictic gestures showed that bilingual adults use deictic gestures more often when speaking their weaker language than when speaking their stronger language (Gullberg, 1998, Hadar et al. 2001, Nicoladis et al., 2007, Sherman & Nicoladis, 2004), thus suggesting a compensatory role for deictic gestures. Research on iconic gestures, on the other hand, presented somewhat inconclusive results. Some studies (e.g., Nicoladis et al., 2007) showed that bilingual adults use more iconic gestures when speaking their weaker language, while others (e.g., Gullberg) showed that bilingual adults produced more iconic gestures when speaking their stronger language.

Some of the more recent work with bilinguals focused on the patterns of gesture use during L2 production when speakers were describing a domain that shows strong cross-linguistic variability in L1 and L2 speech. They found that L2 learners’ gestures continued to show L1-like patterns both in the types of meanings conveyed in gesture and the timing of gesture with speech (Choi & Lantolf, 2008; Stam, 2006); while other studies highlighted the importance of greater proficiency in L2 as an important factor in shifting bilingual speakers toward a more L2-like gesture pattern (Özyürek, 2002).

In summary, research on bilingual gesture is still sparse, and presents somewhat inconclusive results. Therefore, a close examination of the amount and patterns of gestures L2 speakers produce when speaking L2 can provide valuable insight into our understanding of the ‘product’ and ‘process’ of second language acquisition.

In this study, we asked whether and how gestures of bilingual speakers differ from monolingual speakers of each language and whether bilingual speakers rely on gesture more to possibly compensate for the difficulties they encounter in the spoken modality. We focused our analysis on speakers’ descriptions of motion (e.g., girl flips over the beam), which were known to evoke high rates of spontaneous gesture production in native speakers of English and Turkish. We explored the speech and gestures bilingual speakers of English and Turkish produce when speaking L1 and L2, and compared their production to monolingual speakers’ speech and gesture use in each language. Our goal was to determine (1) whether bilinguals would use gesture at higher rates in speaking L2 than L1 to compensate for their difficulties in L2, and (2) whether bilingual speakers’ gestures in L2 would resemble the gestures of native speakers of the same language. Overall, this study project would allow us to understand the complex interplay between spoken and verbal modalities and the effects of L1 on the production of L2.
2. Methods

Sample

Ten bilingual adult Turkish speakers with English as second language ($M_{age} = 29$; range = 22-44), along with 10 monolingual English speakers ($M_{age} = 21$; range = 22-39) and 10 monolingual Turkish speakers ($M_{age} = 32$; range = 18-43) participated in the study. The data from bilingual speakers and monolingual English speakers were collected in Atlanta; the data from monolingual Turkish speakers were gathered in Istanbul. The bilinguals included advanced L2 learners of English, who have started learning English during secondary school years and who have been residing in the United States for at least four years. There were equal numbers of males and females in each group.

Procedure

Each adult was interviewed individually. Monolingual English and monolingual Turkish speakers were interviewed in one session by a native English speaker and a native Turkish speaker, respectively. Bilingual speakers were interviewed in two sessions: once in English by a native English speaker and once in Turkish by a native Turkish speaker. The order of the two languages was counterbalanced across bilingual participants. During the interview, participants were first introduced to a cartoon character named Adam, who performed the motions in the animations. They were then presented with 16 animated motion scenes—one at a time (see Figure 1), and asked to respond in two different ways. In the first condition, they were asked to describe the animations using words (*speech condition*; ‘Tell me what is happening in this clip’). In the second condition, they were asked to describe the animations using their hands without any speech (*gesture-only condition*; ‘Tell me what is happening in this clip only with your hands, without speaking’). Participants described all animations first in the speech condition and then in the gesture-only condition. The gesture-only condition was included as a control to determine whether the patterns we observe in gesture were driven by the accompanying speech (L1 or L2) or by the modality of the expression itself (i.e., gesture), and was administered only once to each monolingual and bilingual speaker.

![Figure 1. Snapshot of a sample motion animation: crawling into a house](image)

Coding

All speech and gestures describing motion were extracted from the transcripts and segmented into target utterances. In the speech condition, each target utterance consisted of a sentence (i.e., a verb and its associated arguments; e.g., ‘he crawls into the house’, ‘he entered the house in a rush’) and the accompanying gestures. In the gesture-only condition, each target utterance referred to the string of gestures produced to describe each of the scenes.
In this study, we focused only on iconic gestures, which convey meaning by their “iconic” resemblance to the different aspects of the motion they depict (e.g., moving the hand up and down rapidly to convey ‘hopping’, wiggling fingers left to right to convey ‘walking towards’); iconic gestures also constituted the predominant gesture type produced by the participants. We coded each utterance for the type of motion element it conveyed, as either manner-only (e.g., circling hands next to body as if running), path-only (moving index finger forward at neutral speed as if going forward), or manner+path conflated (e.g., circling hands forward as if running forward), separately for speech and gesture. For speech utterances that conveyed manner+path, we further specified whether the utterance consisted of a manner verb with a path satellite (MANNER+SUBPATH; e.g., ‘he ran into the house’) or a path verb with a manner adjunct (PATH+ SUBMANNER; e.g., ‘he entered the house running/rapidly’).

We conducted two types of analyses on the coded data, one focusing on the amount and the other on the patterns of gesture and speech use. First, to determine differences in overall amounts of gesture and speech production in L1 and L2, we computed the total number of gestures and words (i.e., tokens) each bilingual speaker produced, separately for the descriptions in L1 and L2. Second, to determine differences in patterns of speech, we determined the types of motion elements conveyed in each spoken utterance (manner, path), the frequency of each element, and whether each speech utterance conveyed information about a single element (manner-only, path-only) or multiple elements (manner+path). For elements that conveyed both manner+path, we categorized them into two as either manner verb+path satellite (MANNER+SUBPATH) or path verb+manner adjunct (PATH+SUBMANNER) to better capture the crosslinguistic variability in the lexicalization patterns of each language. Next, to determine differences in patterns of gesture, we determined the types of motion elements conveyed in each gesture (manner, path), the frequency of each element, and whether each gesture conveyed information about a single element (manner-only, path-only) or multiple elements (manner+path). Differences between the three groups (monolingual English, monolingual Turkish, bilingual English-Turkish) and differences between productions in L1 and L2 within the bilingual group were analyzed by a set of planned ANOVA comparisons, with either one between subject (group type) or one within-subject (language) factors, separately for the descriptions in the speech condition and the descriptions in the gesture-only condition.

3. Results

Do bilingual speakers differ in their amount of gesture use when speaking L1 vs. L2?

Figure 2 shows the overall amounts of gesture and speech production of bilingual speakers when describing the motion scenes in English (L2) and in Turkish (L1). As the Figure illustrates, bilingual speakers used significantly more gestures when speaking L2 than when speaking L1 (M_{12}= 23.90 [SD=17.80] vs. M_{11}= 8.20 [SD=6.09], F(1, 9)= 8.92, p=.01). Interestingly, they also produced more speech (i.e., word tokens) when providing descriptions in L2 as compared to L1 (M_{12}= 220.50 [SD=101.02] vs. M_{11}= 153.60 [SD=68.13]; F(1, 9)= 23.35, p=.001), suggesting that greater gesture production might simply be an outcome of greater speech production. However, even after controlling for the amount of speech, we found that bilingual speakers still showed reliably greater proportion of
gesture use in L2 than in L1 (F(1, 9)= 6.69, p=.03, providing further evidence of greater gesture production in L2.

**A. Speech**

**B. Gesture**

![Figure 2. Mean number of word tokens (Panel A) and gesture tokens (Panel B) produced by bilingual speakers of English and Turkish in L1 (Turkish, gray bars) and in L2 (English, black bars)](image)

**Do monolingual and bilingual speakers differ in their patterns of speech about motion?**

Monolingual English and Turkish speakers differed reliably in their speech about motion, (F(1,18)=7.88, p=.01). Beginning with monolingual speakers, we found that the majority of the verbal descriptions produced by English and Turkish monolinguals (86% and 76%, respectively) conveyed both manner and path information simultaneously. However, as can be seen in Figure 3, the way speakers packaged these two key motion components showed variability between the two languages. English monolinguals exclusively used manner verbs accompanied by path satellites to describe motion scenes (86%; e.g., ‘Adam is flipping over the beam’), while Turkish speakers predominantly (51%) relied on path verbs along with manner adjuncts to describe similar scenes (e.g., ‘Adam kütügün üzerinden bir sıçrayışta geçti’ [Adam passes over the beam in a jump]).

**A. Monolingual speakers**

**B. Bilingual speakers**

![Figure 3. Mean percentage of manner-only, path-only and manner+path descriptions in speech produced by monolingual speakers of English or Turkish (panel A) and bilingual speakers of English and Turkish in L1 or in L2 (panel B)](image)
Turning next to bilingual speakers, we found similar patterns. Bilingual speakers also showed a significant effect of language ($F(1,18)=6.08$, $p=.02$) in their verbal descriptions. Similar to monolingual speakers, the majority of their verbal descriptions conveyed both manner and path information simultaneously, both in English (69%) and in Turkish (72%). Bilingual speakers also followed two different linguistic strategies in packaging manner and path components. As can be seen in Figure 3 (panel B), when describing the scenes in English (L2), they used manner verbs with path satellites; and when describing the same scenes in Turkish (L1), they resorted to the monolingual Turkish pattern, using path verbs accompanied by manner adjuncts. Thus, both monolingual and bilingual speakers expressed manner and path simultaneously in their speech, but using different linguistic strategies appropriate for the language that they are using.

Do monolingual and bilingual speakers differ in their gestures about motion when gestures are accompanied by speech?

We next looked at the patterns of gesture that accompanied motion descriptions in the two languages. Beginning with monolingual speakers, we found that the speakers in the two languages showed reliable cross-linguistic differences in their gesture patterns ($F(1,18)=6.39$, $p=.02$; see Figure 4, Panel A). English speakers frequently synthesized manner and path components into a single gesture (53%), while Turkish speakers relied predominantly on separated gestures, conveying either manner (40%) or path (44%) information.

Bilingual speakers, on the other hand showed no effect of language in their gesture patterns when describing the scenes ($F(1, 18)=1.0$, ns). As Figure 4 (Panel B) illustrates, when describing the scenes in Turkish, they followed the Turkish monolingual pattern by producing separated gestures for either manner (39%) or path (50%) of motion. Interestingly, they continued to show the same L1 pattern when speaking L2. That is, they continued to use predominantly separated gestures (manner-only, path-only), and produced very few gestures (11%) that conveyed manner and path information simultaneously when describing the scenes in English.

A. Monolingual speakers

B. Bilingual speakers

![Figure 4. Mean percentage of manner-only, path-only and manner+path conflated gestures produced by monolingual speakers of English or Turkish (Panel A) and bilingual speakers of English and Turkish in L1 or L2 (Panel B)](image-url)
Do monolingual and bilingual speakers differ in their gestures about motion when gestures are produced without speech?

Next we asked whether the patterns that we observed in gesture were simply driven by speech that accompanied the gesture or by the gestures themselves. First looking at the patterns in gesture use in monolingual speakers, we found that both English and Turkish speakers showed the same gesture patterns, predominantly using manner-path conflated gestures (77%-93%), with no reliable differences (F(1,18)=1, ns; see Figure 5, Panel A). The patterns were exactly the same in bilingual speakers, with a predominance of manner-path conflated gestures (85%) in the gesture-only condition. That is, bilingual speakers did not differ reliably from either English monolinguals (F(1, 18)=3.24, ns) or Turkish monolinguals (F(1, 18)=2.25, ns), when asked to describe the scenes without using speech.

A. Monolingual speakers

B. Bilingual speakers

Figure 5. Mean percentage of manner-only, path-only and manner+path conflated gestures produced by monolingual speakers of English or Turkish (Panel A) and bilingual speakers of English and Turkish (Panel B)

4. Conclusions

Gestures constitute part of what learners acquire in a new language and show variability across different languages. But how closely is gesture tied to the language one speaks? We explored this question by studying the speech and gestures produced by 10 adult advanced second language learners of English (Turkish as L1) in comparison to 10 adult monolingual English and 10 adult monolingual Turkish speakers. We focused on the expression of physical motion, and asked speakers to describe motion scenes first with their words (speech condition) and then with their hands (gesture-only condition). We found strong crosslinguistic differences in monolingual and bilingual speakers’ speech and gesture patterns in the speech condition, but close crosslinguistic similarities in their gestures in the gesture-only condition. We also found that bilingual speakers continued to show L1 patterns in gesture, while describing the motion scenes in L2—even if they showed strong L2 patterns in speech. Our findings suggest that acquisition of native-like gesture patterns takes longer to establish than acquisition of native-like speech patterns, and gestures follow language-specific patterns only when accompanied by speech.
Our findings also extend earlier work on gesture development in first language acquisition to the domain of second language acquisition. This earlier work (Özyurek et al, 2008) showed that children learning English or Turkish as their native language begin by producing separate gestures for manner and path at age 3 when describing motion. These early patterns change into adult-like language-specific patterns in each language around age 9, with children learning English synthesizing manner and path components into a single gesture and children learning Turkish continuing to produce separate gestures for each motion component. These findings thus suggest that acquisition of language-specific gestures patterns takes longer to establish than acquisition of native speech patterns when children are learning their first language (see also Özçalışkan & Goldin-Meadow, 2011; Özçalışkan et al., 2012 for a different pattern of findings in first language development in English). Similarly, the bilingual speakers in our study were not yet producing native gesture patterns unique to English when speaking English, even if they were already in full mastery of the language-specific speech patterns of their second language.

Overall, our results suggest that acquisition of native gesture patterns might serve as a better index of native-like fluency in second language than evidence of this fluency in speech alone.

5. References


